

Regular Expressions (RegEx)

Regular expression is like another language

- What is a regular expression?
- Literal (or normal characters)
 - Alphanumeric
 - abc...ABC...0123...
 - Punctuation
 - `-_.,:;=()/+ *%&{}[]? !$'^|\<>"@#`
- Just like languages Regular expressions also have dialects
 - awk, egrep, Emacs, grep, Perl, POSIX, Tcl, PROSITE

Why use a regular expression?

ALSTDNVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIIY
GFPEPTRAHCDFEPVFRQESCFYWLTGVNEADCAYFLDIETG
KEILFYDPDIPQAYIIWFGELATIDDIKKKYGFEDVRLMPKIQETL
AEYKLLKIHITLPETCILKGYDALLDKNEFIDVVGELRQIKDDDE
MVLIQYACDVNSFAVRDTFKKVHPKMWEHQVEANLIKHYV
DYCRCFAFSTIVCSGENCSILHYHHNNKFIEDGELILIDTGCE
YNCAADNTRTIPANGKFSPDQRAVYQAVLDCHNYVVAHAK
PGVWPDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALA
VFYPHGLGHGMGIDCHEIAGWPRGTCTRGKKPHHSFVRFGR
TLEKGVVITNEPGCYFIRPSYNAAFADPEKSKYINKEVCERLRK
TVGGVRIEDDLLITEDGCKVLSNIPKEIDEIEAFMAKKE

Why use a regular expression?

ALSTDNVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIIY
GFPEPTRAHCDFEPVFRQESCFYWLTGVNEADCAYFLDIETG
KEILFYDPDIPQAYIIWFGELATIDDIKKKYGFEDVRLMPKIQETL
AEYKLLKIHITLPETCILKGYDALLDKNEFIDVVGELRQIKDDDE
MVLIQYACDVNSFAVRDTFKKVHPKMWEHQVEANLIKHYV
DYCRCFAFSTIVCSGENCSILHYHHNNKFIEDGELILIDTGCE
YNCAADNTRTIPANGKFSPDQRAVYQAVLDCHNYVVAHAK
PGVWPDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALA
VFYPHGLGHGMGIDCHEIAGWPRGTCTRGKKPHHSFVRFGR
TLEKGVVITNEPGCYFIRPSYNAAFADPEKSKYINKEVCERLRK
TVGGVRIEDDLLITEDGCKVLSNIPKEIDEIEAFMAKKE

What does a regular expression look like?

- `ALSTDNVANRPMPKPEMF....`
- Text: The sequence must start with an alanine, followed by any amino acid, followed by a serine or a threonine, two times, followed by any amino acid or nothing, followed by any amino acid except a valine.
- Regex: `^A.[ST]{2}.?[^\V]`